

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:08:59 ; Search time 299.73 Seconds
(without alignments)
16.815 Million cell updates/sec

Title: US-09-331-631a-1_COPY_74_116
Perfect score: 248
Sequence: 1 NQEDPTECCQCCRCRCRQOE.....RQOYCORCKEICEEEY 43

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPTRMBL_15:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	248	100.0	666	10 Q9SP15	Q9SP15 macadamia i
2	242	97.6	666	10 Q9SP14	Q9SP14 macadamia i
3	241	97.2	625	10 Q9SP13	Q9SP13 macadamia i
4	119	48.0	593	10 Q9SEW4	Q9SEW4 juglans reg
5	117	47.2	525	10 Q43358	Q43358 theobroma c
6	96	38.7	411	5 P91419	P91419 caenorhabdi
7	80.5	32.5	425	5 Q17400	Q17400 caenorhabdi
8	80.5	32.5	600	5 Q17401	Q17401 caenorhabdi
9	80.5	32.5	810	10 Q92W13	Q92W13 cucurbita m
10	79	31.9	572	5 Q19594	Q19594 caenorhabdi
11	77.5	31.2	242	5 Q19919	Q19919 caenorhabdi
12	75	30.2	339	5 Q9NAR1	Q9NAR1 caenorhabdi
13	74	29.8	388	5 Q16500	Q16500 caenorhabdi
14	74	29.8	388	5 Q16501	Q16501 caenorhabdi
15	74	29.8	438	5 Q16502	Q16502 caenorhabdi
16	74	29.8	445	5 Q16511	Q16511 caenorhabdi
17	73.5	29.6	335	5 Q9NF70	Q9NF70 caenorhabdi
18	73.5	29.6	335	5 Q9NF69	Q9NF69 caenorhabdi
19	73.5	29.6	419	5 Q9NA38	Q9NA38 caenorhabdi

20	73.5	29.6	420	5 Q9NA61	Q9NA61 caenorhabdi
21	73	29.4	388	5 Q44606	Q44606 caenorhabdi
22	72.5	29.2	330	5 Q18118	Q18118 caenorhabdi
23	72.5	29.2	709	5 Q9NAR2	Q9NAR2 caenorhabdi
24	71.5	28.8	335	5 Q9N6L3	Q9N6L3 caenorhabdi
25	70.5	28.4	378	5 Q27383	Q27383 caenorhabdi
26	70	28.2	1513	5 Q17970	Q17970 caenorhabdi
27	68.5	27.6	393	10 Q9ZTP0	Q9ZTP0 oryza sativ
28	68.5	27.6	654	5 Q17982	Q17982 caenorhabdi
29	68	27.4	594	5 Q77337	Q77337 plasmodium
30	66.5	26.8	273	5 Q45362	Q45362 caenorhabdi
31	66.5	26.8	314	5 Q23390	Q23390 caenorhabdi
32	66.5	26.8	356	5 Q17316	Q17316 ceratilis c
33	66.5	26.8	637	10 Q03678	Q03678 hordeum vul
34	65	26.2	111	5 Q19054	Q19054 caenorhabdi
35	65	26.2	539	4 Q9NUA2	Q9NUA2 homo sapien
36	65	26.2	1306	5 Q77273	Q77273 drosophila
37	63	25.4	243	5 Q9NL90	Q9NL90 entamoeba d
38	63	25.4	910	11 Q54899	Q54899 mus musculu
39	63	25.4	910	11 Q88704	Q88704 mus musculu
40	62.5	25.2	342	4 Q9UK28	Q9UK28 homo sapien
41	62	25.0	425	5 Q15755	Q15755 dictyosteli
42	62	25.0	648	5 Q9TXB8	Q9TXB8 dictyosteli
43	62	25.0	715	4 Q9UEX4	Q9UEX4 homo sapien
44	62	25.0	1080	5 Q9NZM8	Q9NZM8 drosophila
45	61	24.6	400	6 Q95337	Q95337 tupiaa glis

ALIGNMENTS

RESULT 1
Q9SP15 PRELIMINARY; PRT; 666 AA.
AC Q9SP15;
DT 01-MAY-2000 (Tremblrel, 13, Created)
DT 01-MAY-2000 (Tremblrel, 13, Last sequence update)
DT 01-OCT-2000 (Tremblrel, 15, Last annotation update)
DE VICILIN PRECURSOR.
GN AMP2.
OS Macadamia integrifolia (Macadamia nut).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.
OC NCBI_TaxID=60698;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NUIT KERNEL;
RA Marcus J.P., Goulter K.C., Green J.L., Manners J.M.;
RT "A family of antimicrobial peptides is produced by processing of a 7S
RT globulin protein in Macadamia integrifolia kernels.";
RL Plant J. 0:0-0(1999).
DR EMBL: AF161883; AAD54244.1; -
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR001113; -
DR PFWM: PF00546; Seedstore-7s; 1.
SQ SEQUENCE 666 AA; 78217 MW; C75228884B2DF0224 CRC64;

Query Match 100.0%; Score 248; DB 10; Length 666;
Best Local Similarity 100.0%; Pred. No. 1.2e-22;
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 NQEDPTECCQCCRCRCRQOESGPRQOYCORCKEICEEEY 43
DB 74 NQEDPTECCQCCRCRCRQOESGPRQOYCORCKEICEEEY 116

RESULT 2
Q9SP14 PRELIMINARY; PRT; 666 AA.
AC Q9SP14;
DT 01-MAY-2000 (Tremblrel, 13, Created)
DT 01-MAY-2000 (Tremblrel, 13, Last sequence update)

DT 01-OCT-2000 (TREMBLeref. 15, Last annotation update)
DE VICILIN PRECURSOR.
GN AMP2
OS Macadamia integrifolia (Macadamia nut).
OC Eucaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Proteaceae; Macadamia.
OX NCBI_taxID=60698;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NUT KERNEL;
RA Marcus J.P., Coulter K.C., Green J.L., Manners J.M.;
RT "A family of antimicrobial peptides is produced by processing of a 7S
RT globulin protein in Macadamia integrifolia.";
RL Plant J. 0:0-0(1999).
DR EMBL; AF161884; AAD54245.1; -
DR HSSP; P02853; 2PHL.
DR INTERPRO: IPR001113; -
DR Pfam: PF00546; Seedstore_7s; 1.
SO SEQUENCE 666 AA; 78243 MW; 10ECA22EF810F8A7B CRC64;

	Query Match	Score	DB	Length
	Best Local Similarity	95.3%	Pred. No. 6, 5e-22;	
	Matches	41;	Conservative	2; Mismatches 0; Indels 0; Gaps 0;
QY	1	NOEDPOTECOCQRCRCROQESGPRQDQTCQRCRCCTCEEEERY	43	
DB	74	NODDPOTDCQRCRCROQESGPRQDQTCQRCRCCTCEEEERY	116	

Q	SEQUENCE	625 AA:	73586 MW:	415808A89D370296 CIRC64:
09SP_L3	1			
09SP_L3	PRELIMINARY;	PRT;	625 AA.	
AC	09SP_L3;			
DT	01-MAY-2000 (TReMBLrel. 13, Created)			
DT	01-MAY-2000 (TReMBLrel. 13, Last sequence update)			
DT	01-OCT-2000 (TReMBLrel. 15, Last annotation update)			
DE	VICILIN PRECURSOR (FRAGMENT).			
GN	AMP2.			
OS	Macadamia integrifolia (Macadamia nut).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Manoliophyta; eudicotyledons; Proteaceae; Macadamia.			
OX	NCBI_TaxID=60698;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=NUT KERNEL;			
RA	Marcus J.P., Goulter K.C., Green J.L., Manners J.M.;			
RT	"A family of antimicrobial peptides is produced by processing of a 79S			
RT	globulin protein in Macadamia integrifolia.";			
RL	Plant J. 0:0-0(1999).			
DR	EMBL: AF161885; AAD54246.1; -.			
DR	HSSP: P02853; 2PHL.			
DR	INTERPRO: IPR001113; -.			
DR	PFAM: PF00546; Seedstore_7s; 1.			
FT	NON_TER			
Q	1			

[illegible]

RESULT	4	
Q9SEW4		
ID	Q9SEW4	
AC	Q9SEW4	
DT	01-MAY-2000 (TREMBlrel. 13, Created)	
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)	
	PRELIMINARY;	PRT; 593 AA.

DT 01-OCT2000 (TEMBLRELE. 15, last annotation update)
DE VICILIN-LIKE PROTEIN PRECURSOR (FRAGMENT).
OS Juglans regia (English walnut).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
CC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Fagales; Juglandaceae; Juglans.
OX NCBI_TaxID=51240;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SUNLAND; TISSUE=SOMATIC EMBRYO LINE;
RA Tenber S.S., Jarvis K.C., Peterson W.R., Dandekar A.M., Ansari A.A.
RT Identification and cloning of a cDNA encoding a vicilin-like protein
RT Jug r 2, from English walnut kernel (Juglans regia): a major food
RT allergen.
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF066055; AAF18269.1; -.
DR HSSP; P02853; 2PHT.
DR INTERPRO; IPR001113; -.
DR Pfam; PF00546; Seedstore_7s; 1.
FT NON_TER 1
SQ SEQUENCE 593 AA; 65990 MW; 9BA127E19A18C0D8 CRC64;

	Query Match	48.0%	Score 119;	DB 10;	Length 593;
	Best Local Similarity	51.2%;	Pred. No. 5,3e-07;		
	Matches	21; Conservative	7; Mismatches	13;	Indels
					Gaps 0;
Oy	1 NOEDPQTECCQCRCRGRROESPRROOYCGRRCKETCEE	41			
Dd	10 NPDREIYROCQEYCRGGGGRKQDCCQLICEERLEDDQ	50			

RESULT	5			
043358				
ID	043358	PRELIMINARY;	PRT;	525 AA.
AC	043358;			
DT	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, last sequence update)			
DT	01-OCT-2000 (TREMBLrel. 15, last annotation update)			
DE	VICILIN PRECURSOR.			
GN	CSV.			
OS	Theobroma cacao (Cacao).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;			
OC	Malvales; Malvaceae; Theobroma.			
OX	NCBI_TaxID=3641;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LEAVES;			
RA	MEDLINE=92288309; PubMed=1600151;			
RT	McHenry L., Fritz P.D.;			
RT	"Comparison of the structure and nucleotide sequences of vicilin genes			
RL	of cocoa and cotton raise questions about vicilin evolution.";			
RL	Plant Mol. Biol. 18:1173-1176(1992).			
DR	EMBL: X62625; CAA44493.1;			
DR	EMBL: X62626; CAA44494.1.;			
DR	HSSP; P02853; 2PRL.			
DR	MENDEL; 30919; Thecc;1186;30919.			
DR	INTERPRO: IPR001113.;			
DR	PFAM: PF00546; seedstore_7s; 1.			
DR	PRODOM: PD081059.;			
DR	PRODOM: PD081059.;			
KW	Signal.			
FT	CHAIN	1	24	POTENTIAL.
FT	CHAIN	25	525	VICILIN.
SQ	SEQUENCE	525 AA;	60796 MW;	19114CDSC248905D CRC64;

Query Match	47.2%	Score 117,	DB 10:	Length 525:
Best Local Similarity	50.0%;	Pred. No.	8.4e-07;	
Matches	Conservative 10;	Mismatches	10;	Indels 0; Gaps 0;
OY	2 QEDPTECCGCCRCRQDSGFPDQOYYCQRCKELCEEEF	41		
	: : : : : : : : : : : :			

DT 01-MAY-1999 (TREMblrel. 10, created)
DT 01-MAY-1999 (TREMblrel. 10, last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, last annotation update)
DE PV100.
OS *Cucurbita maxima* (Pumpkin) (Winter squash).
OC Euarysta: Viridiplantae: Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
OC Cucurbitales; Cucurbitaceae; Cucurbita.
OX NCBI_TaxID=3661;
RN [1]
RP
RC STRAIN=KUBOKAWA AMAKURI NANKIN; TISSUE=COTYLEDON;
RX MEDLINE=99107919; PubMed=9891029;
RA Yamada K., Shimada T., Kondo M., Nishimura M., Hara-Nishimura I.;
RT "Multiple functional proteins are produced by cleaving Asn-Gln bonds
of a single precursor by vacuolar processing enzyme.";
RL J. Biol. Chem. 274:2563-2570(1999).
DR EMBL: AB019195; BAA34056.1; "-"
DR HSSP: P02853; 2PHL.
DR INTERPRO: IPR001113; "-"
DR Pfam: PF00546; Seedstore_7s; 1.
DR PRODOM: PD081059; "-1"
SQ SEQUENCE 810 AA; 97314 MW; A829A3F7542266AB CMC64;

[illegible]

RESULT	10			
Q19594				
ID	Q19594	PRELIMINARY:	PRT:	572 AA.
AC	Q19594;			
DT	01-JUN-1998 (Tremblrel. 06, Created)			
DT	01-JUN-1998 (Tremblrel. 06, Last sequence update)			
DT	01-OCT-2000 (Tremblrel. 15, Last annotation update)			
DE	HYPOTHEtical 62.4 KDA PROTEIN F19G12.7 IN CHROMOSOME X PRECURSOR.			
GN	F19G12.7.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditidae;			
OC	Rhabditidae; Pelodierinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BRISTOL N2;			
RA	Nhan M., Le T.,;			
RL	Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.			
CC	-!- SIMILARITY: BELONGS TO FAMILY UPP.			
DR	EMBL: U51997; AAC48159.1; -.			
DR	MORMEP: F19G12.7; CE07090.			
KM	HYPOthetical protein: Signal.			
FT	SIGNAL			
FT	1			
FT	22			
FT	572			
FT	CHAIN			POTENTIAL.
FT	572 AA;			HYPOTHEtical PROTEIN F19G12.7.
FT	62384 MW;			BOA542FDB850AAB9 CRC64;
FT	SEQUENCE			

Query Match	31.9%	Score 79:	DB 5:	Length 572:
Best Local Similarity	34.0%	Pred. No.	0.037:	
Matches 17; Conservative	8;	Mismatches	7;	Indels 18; Gaps 3;

```

OY      8  ECQCCRCRCRQ-----QESGPRQQYCCRRCKE---ICEEE 40
          :||||| |:|      | | | | | | | | | | | | | | | |
Db      381  QCQCQONTCCQGFAPVCEQGCASLCHQPSAPQCQQ--CQNTCCQGFAPVCEQ 429

```

RESULT 11	
Q19919	
ID Q19919	PRELIMINARY; PRT; 242 AA.

AC 019919;
DT 01-JUN-1998 (TReMBLrel_06, Created)
DT 01-JUN-1998 (TReMBLrel_06, last sequence update)
DT 01-JUN-1998 (TReMBLrel_06, last annotation update)
DE HYPOTHEETICAL 26.2 KDA PROTEIN F3IA3.1 IN CHROMOSOME X.
GN F3IA3.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditoidea;
OC Rhabditidae; Peloiderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL NZ;
RA Murray J., De T.T.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBD databases.
CC -1- SIMILARITY: BELONGS TO FAMILY UPF.
DR EMBL; U58742; AAB36856.1; -.
DR WORMPEP; F3IA3.1; CE07158.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 3
FT POTENTIAL 23
SQ SEQUENCE 242 AA; 26213 MW; C7D65966B00829317 CRC64;

[illegible]

	RESULT	12	
Q9N4R1	ID	Q9N4R1	PRELIMINARY; PTM: 339 AA.
Q9N4R1	AC	Q9N4R1	
DT	01-OCT-2000	(TREMBLrel. 15, Created)	
DT	01-OCT-2000	(TREMBLrel. 15, Last sequence update)	
DT	01-OCT-2000	(TREMBLrel. 15, Last annotation update)	
DE	HYPOTHETICAL PROTEIN Y5H2A.C.		
GN	Y5H2A.C.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdillida; Rhabditoidea;		
OC	Rhabdillidae; Pelodierinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BRISTOL N2;		
RX	MEDLINE=99069613; PubMed=9851916;		
RA	None;		
RT	"Genome sequence of the nematode C. elegans: a platform for		
RT	investigating biology. The C. elegans Sequencing Consortium."		
RL	Science 282:2012-2018(1998).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=BRISTOL N2;		
RA	Waterston R.;		
RL	Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.		
EMBL	AC006809; AAF59626.1; -		
SQ	SEQUENCE 339 AA; 36218 MW; 40233423237C432D CRC64;		

Query Match	30.2%	Score	75	DB	5	Length	339
Best Local Similarity	29.1%	Pred. No.	0.075				
Matches	16	Conservative	11	Mismatches	12	Indels	16
						Gaps	3

```
Qy      2 QEDPQTEGQCQQRK-----QESGPR-----QQYCORCKE---ICEEE 40
      1:      ||||| 1:      || 11 1:      ::      ::
Db      167 QQASAPQCQCQCQNNCCQGFAPVCQQQCAPQCTISSAPQCQCQCCTTCQGFAPVCQQQ 221
```

RESULT
016500

Fri Mar 2 09:28:55 2001

us-09-331-631a-1_copy_74_116.rspt

Page 6

	Query Match	29.6%	Score 74;	DB 5;	length 438;
	Best Local Similarity	30.2%	Pred No. 0.1%;		
	Matches 13;	Conservative 12;	Mismatches 12;	Indels 6;	Gaps 2;
OY	3	EDPOTECO-CORRC-----ROESGPRQOXYCORRKEICEE	39		
		: :: :: :: :			
		:: :: :: :: :			
Dd	191	QAAGACGCPCCCGCQCSCSVCVGGQQDQSTGCDFACMTCGSDTICQQ	233		

Search completed: March 1, 2001, 16:09:02
Job time: 1546 sec

-1_copy-74-116.rsp